Environment-wide studies (EWAS) on Metabolic-syndrome-related phenotypes: T2D and Cholesterol

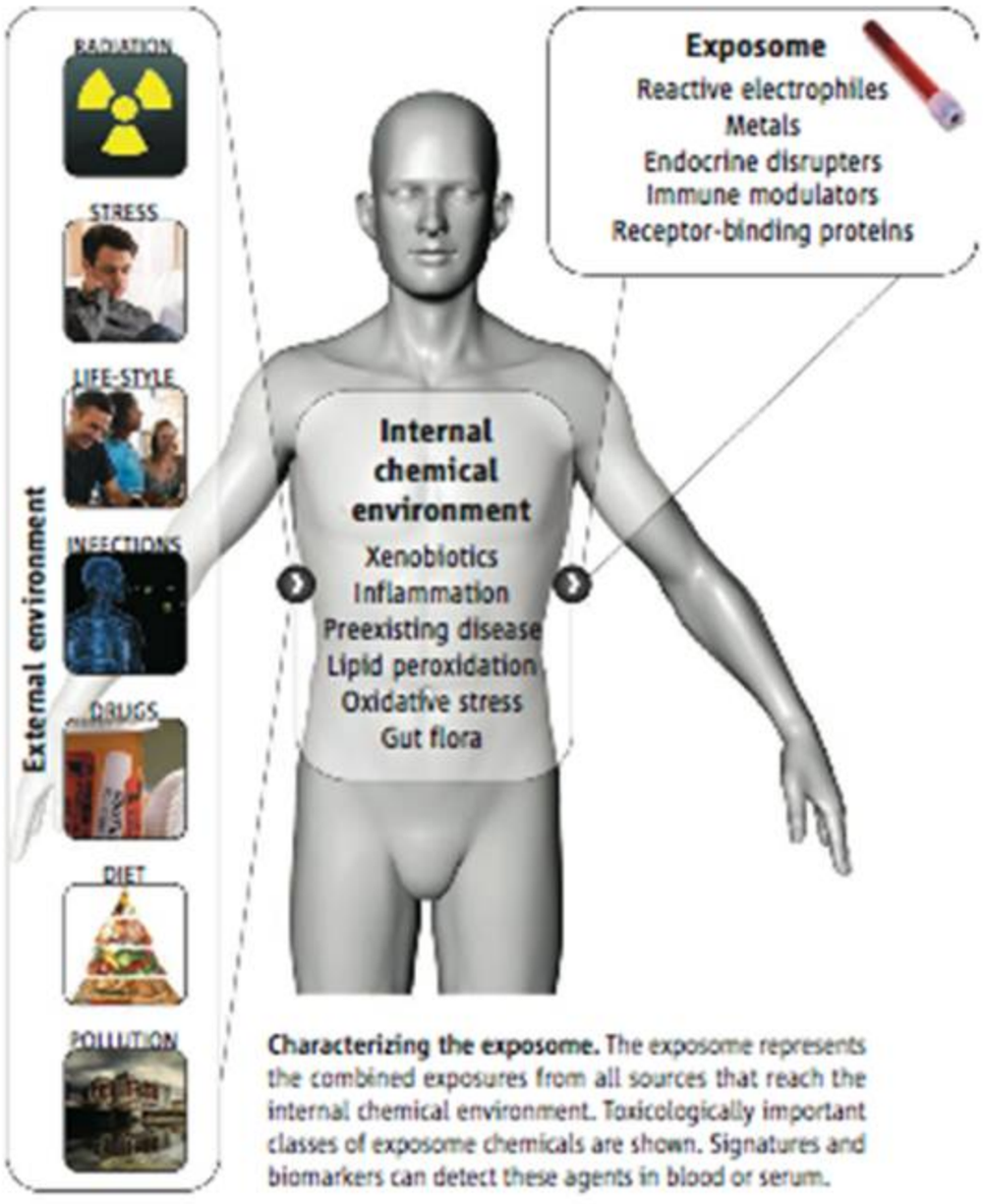
Chirag Patel, MS
Program in Biomedical Informatics
Department of Medicine and Pediatrics
NTP Workshop, 1/11/11

Butte Lab
Stanford Center for Biomedical Informatics Research

“Exposome”-wide measures on an epidemiological scale

What if we could define and use *exposome* as we do the *genome*?

ie, GWAS to “*E*”WAS



Characterizing the exposome. The exposome represents the combined exposures from all sources that reach the internal chemical environment. Toxicologically important classes of exposome chemicals are shown. Signatures and biomarkers can detect these agents in blood or serum.

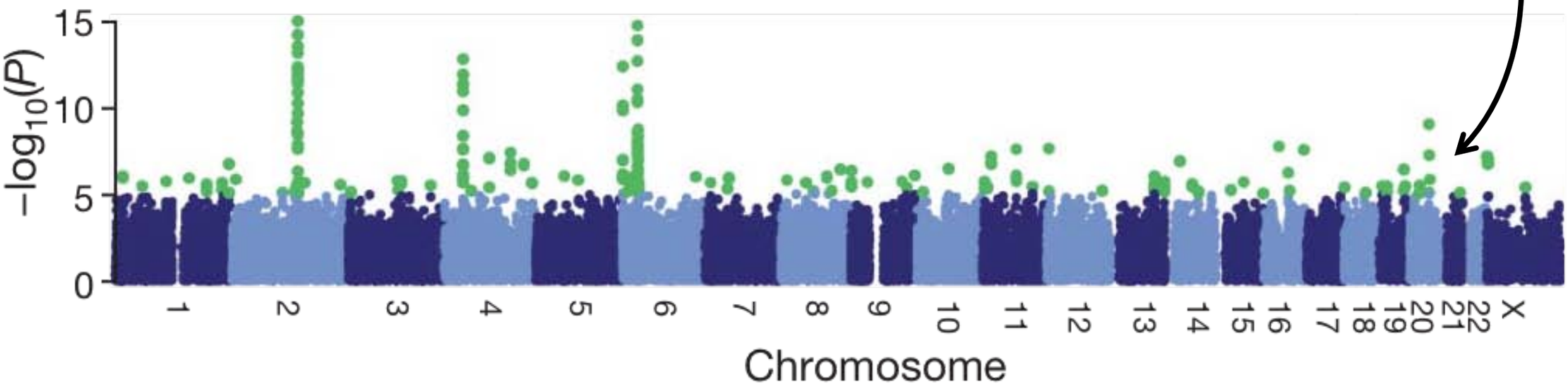
HapMap project:
http://hapmap.ncbi.nlm.nih.gov
/

Illumina “variant chip”

Rappaport S, Smyth M. Environment and Disease Risks.
Science (2010) vol. 330 (6003) pp. 460-461

Genome-Wide Association Studies (GWAS)

	AA	Aa	aa
case			
control			

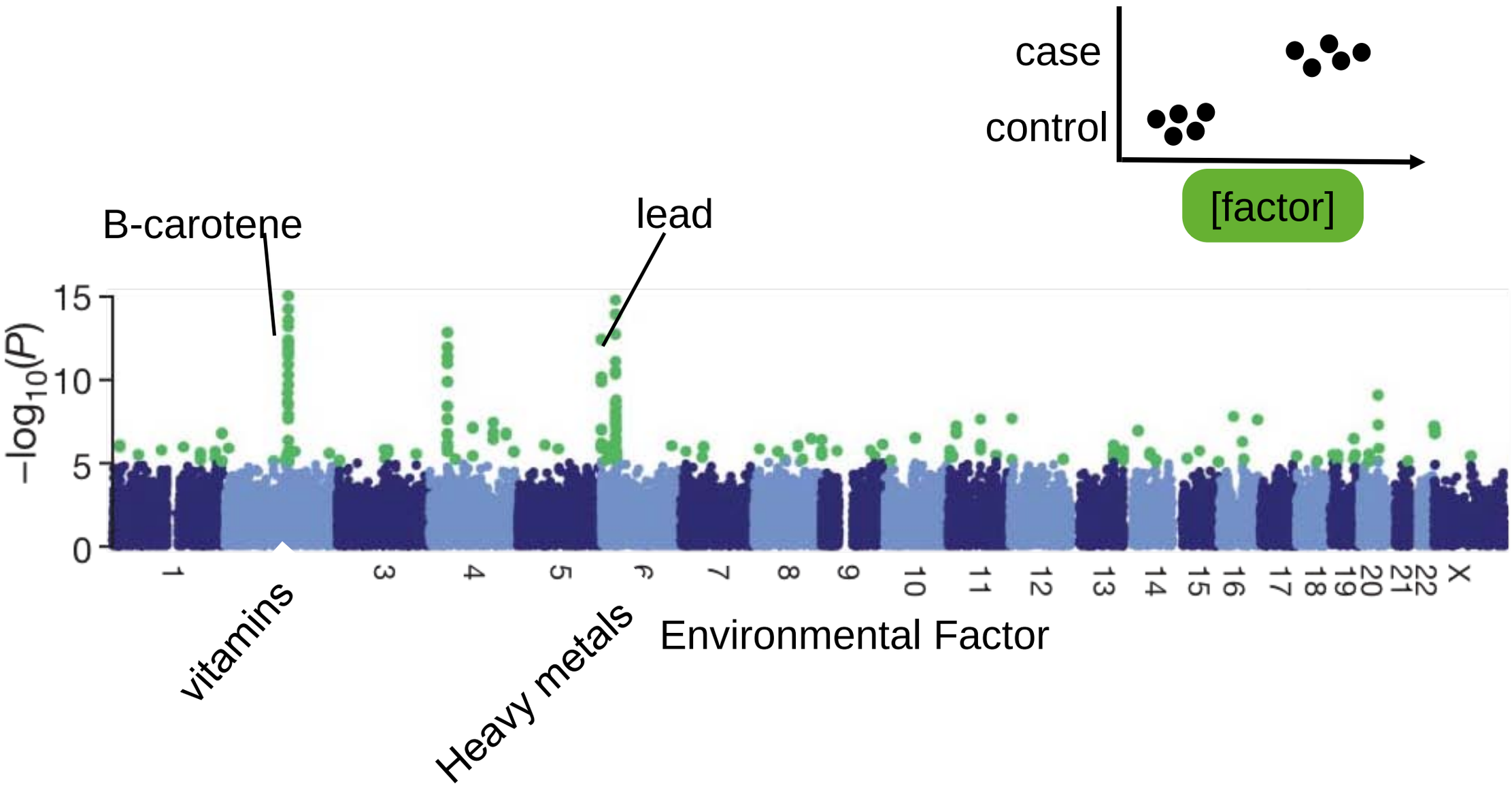


~100,000 - 1,000,000 association tests

What specific genetic loci are associated to disease?

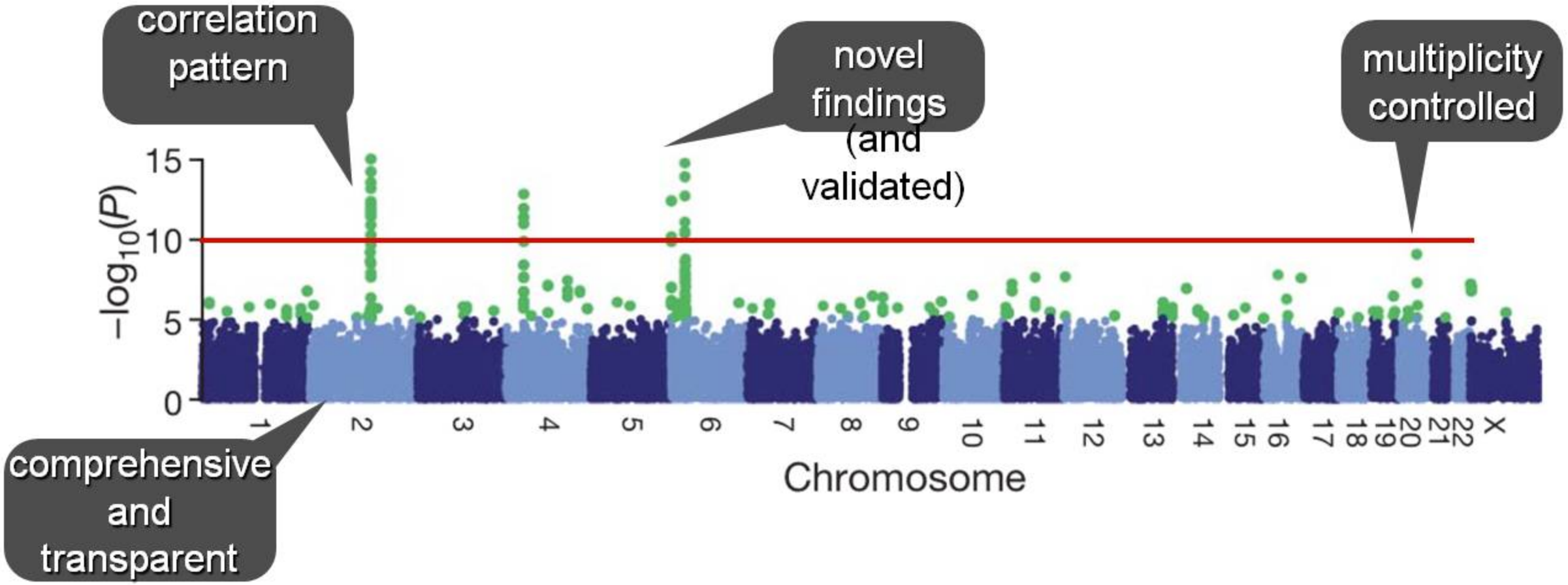
WTCCC, 2007

Environment-Wide Association Studies (EWAS)



What specific **environmental** “loci” are associated to disease?:
ie, T2D, lipid levels, obesity, etc?

Why “EWAS”?



Potential for agnostic, comprehensive, and systematic association studies

However: confounding & reverse causality bias are multiplied

Ioannidis JPAI, Loy EY, et al, (2009) Researching genetic vs nongenetic determinants of diseases: a comparison and proposed unification. *Sci Trans Med* vol. 1(7)7ps8

NHANES Measurements
Environmental "E-Loci" Chip

Demographics (N ~=10,000)	Examination (N ~= 3000)
Age Sex Income Education Ethnicity	Blood Pressure Body Measurement Diet & Nutrient Intake Vision Oral Health
Laboratory (N ~= 3000)	Questionnaire (N ~=10,000)
Biochemistry: Triglycerides, cholesterol, glucose, Cl, Na, K, PSA, etc. **Exposures**: Heavy metals, dioxins, PCBs, phenols, phthalates; Infectious Diseases Allergens	Disease & Health Status Drug use Physical Activity Health & Fitness History Occupation

Cohorts:
1999-2000
2001-2002
2003-2004
2005-2006

Restricted-use: genetic variants!

EWAS Methodology

Disease status Classification

B Disease Phenotypes

Fasting Glucose
> 125 mg/dL?
N=109-3190
(8% of total)

log10(trigly.)
N=109-3618

log10(LDL)
N=101-3368

log10(HDL)
N=222-7485

A Number of Factors Per Class

Environmental Factor "Classes"

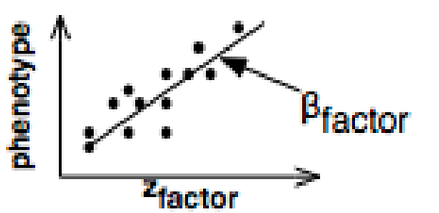
	1999-2000	2001-2002	2003-2004	2005-2006
Acrylamide	0	0	2	0
Allergen Test	0	0	0	20
Bacterial	8	13	17	1
Cotinine	1	1	1	1
Diakyl	7	7	6	0
Dioxins	5	7	7	0
Furans	5	5	9	0
Heavy Metals	18	18	23	25
Hydrocarbons	14	22	21	0
Latex	1	0	0	0
Carotenoid Nutrients	0	6	15	7
Mineral Nutrients	2	2	2	1
Vitamin A	3	3	3	3
Vitamin B	4	4	5	3
Vitamin C	0	0	1	1
Vitamin D	0	1	1	1
Vitamin E	2	2	3	2
Polychlorinated Biphenyls	23	26	38	0
Perchlorate	0	0	2	0
Pesticides, Atrazine	0	0	5	0
Pesticides, Carbamate	0	0	1	0
Pesticides, Chlorophenol	0	0	1	1
Pesticides, Organochlorine	10	13	11	0
Pesticides, Organophosphate	2	2	2	0
Pesticides, Pyrethyroid	1	1	1	0
Phenols	15	11	9	12
Phthalates	7	12	12	0
Phytoestrogens	6	6	6	0
Polybrominated Ethers	0	0	12	0
Polyflourochemicals	0	0	10	12
Virus	6	6	10	6
Volatile Compounds	29	14	22	0
total	**169**	**182**	**258**	**96**

Factor "class"

Linear Modeling

C Screening For Disease-associated Factors

z_{factor} = transformed x_{factor}
- age, age*age, sex, ethnicity, BMI, SES



Empirical False Discovery Estimation

D Validation

Empirical FDR estimation
Permute Phenotype Levels
1000x

FDR(α) ≤ 10%
P-value(β_{factor}) < α in 2 or more cohorts?

"Validation"

EWAS Methodology, cont'd

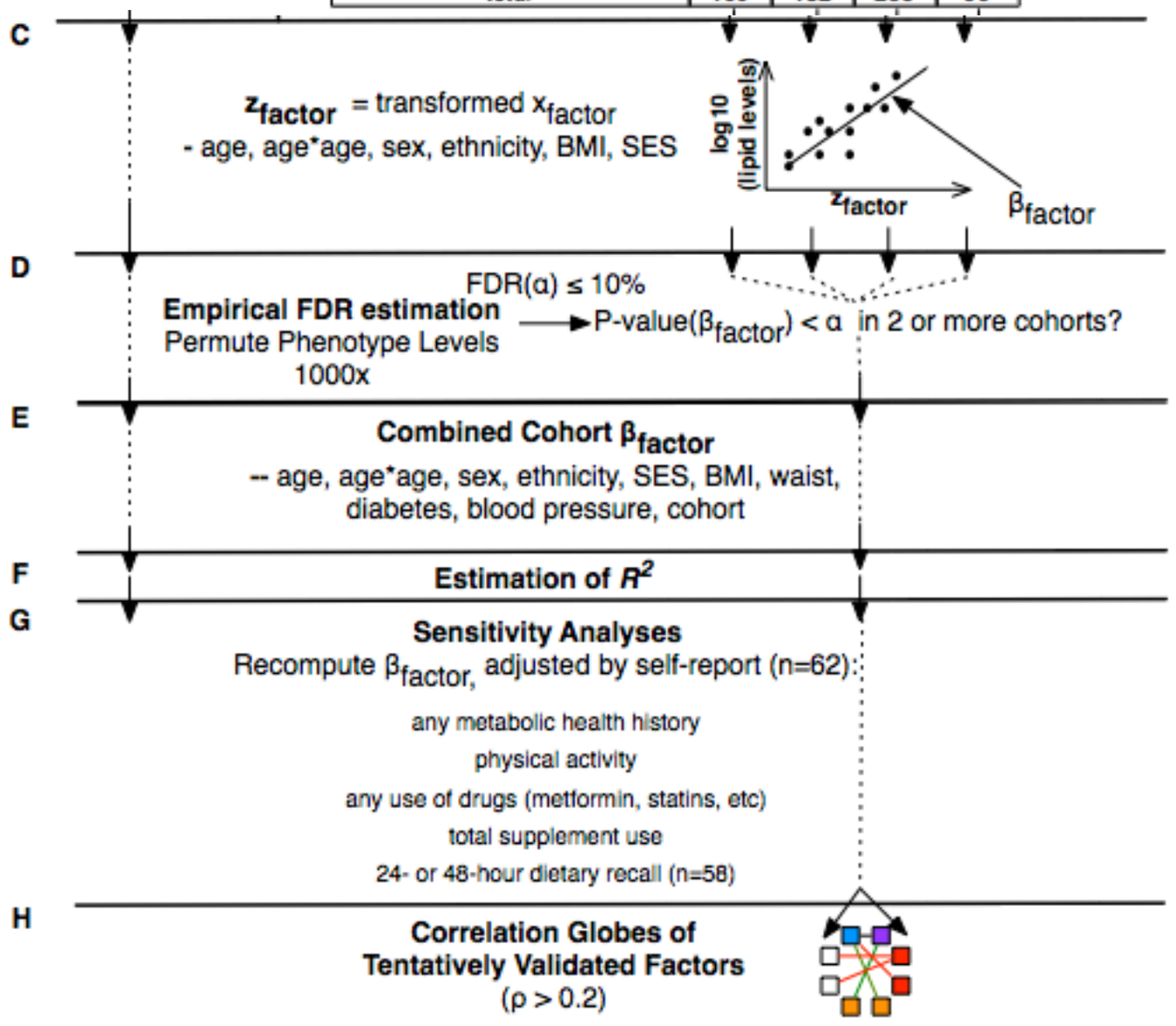
Linear Modeling

Empirical False Discovery Estimation

Sensitivity analyses: adjust for additional factors

Correlation globes

Pesticides, Chlorophenol	0	0	1	1
Pesticides, Organochlorine	10	13	11	0
Pesticides, Organophosphate	2	2	2	0
Pesticides, Pyrethyroid	1	1	1	0
Phenols	15	11	9	12
Phthalates	7	12	12	0
Phytoestrogens	6	6	6	0
Polybrominated Ethers	0	0	12	0
Polyflourochemicals	0	0	10	12
Virus	6	6	10	6
Volatile Compounds	29	14	22	0
total	**169**	**182**	**258**	**96**



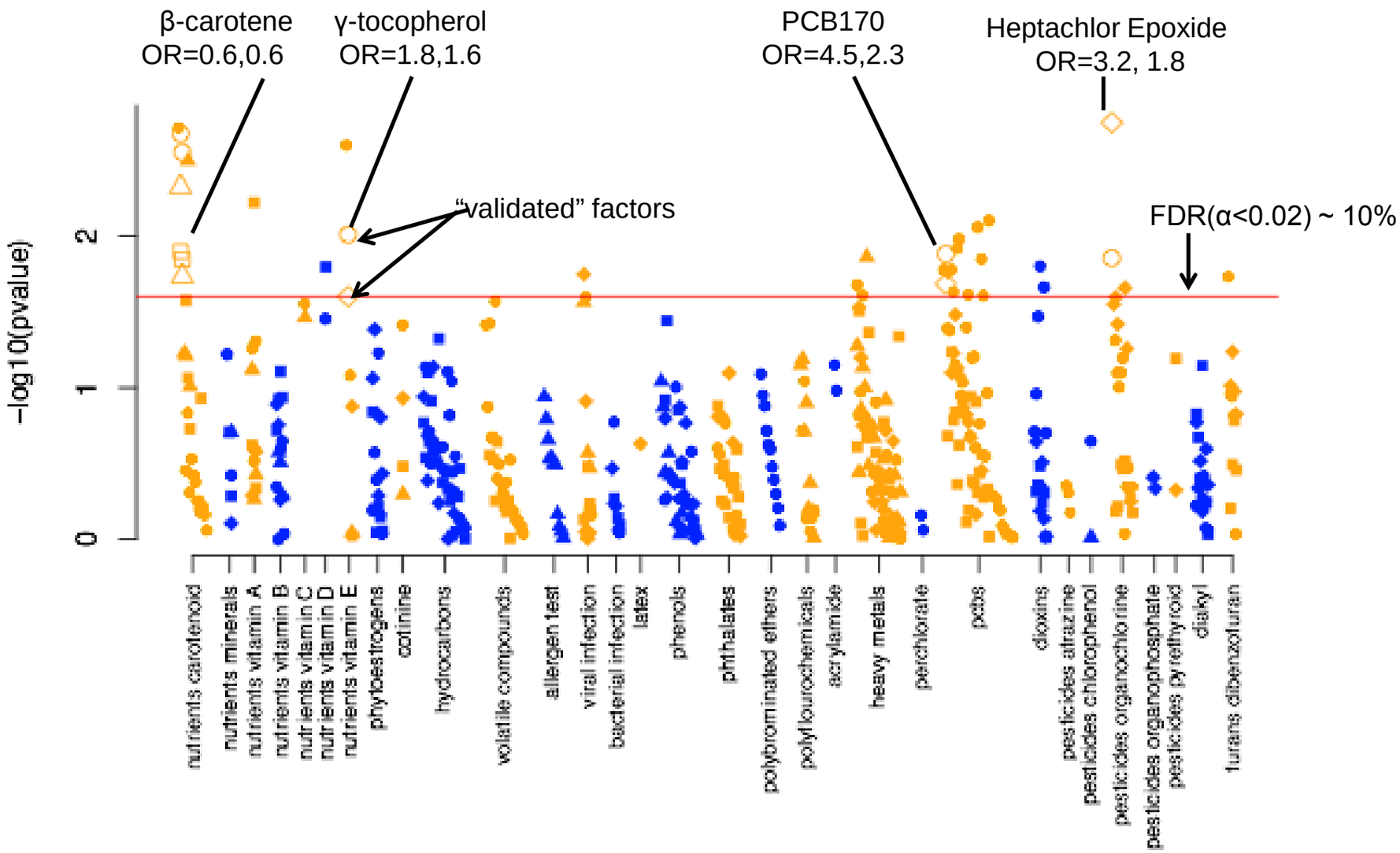
EWAS on T2DM

cohort markers
1999-2000 ◆
2001-2002 ■
2003-2004 ●
2005-2006 ▲

<u>"Novel" Findings:</u>
heptachlor
γ-tocopherol

<u>Known Associations:</u>
β-carotene
vitamin D
PCBs

<u>Interesting Patterns:</u>
dioxins, pesticides, PCBs

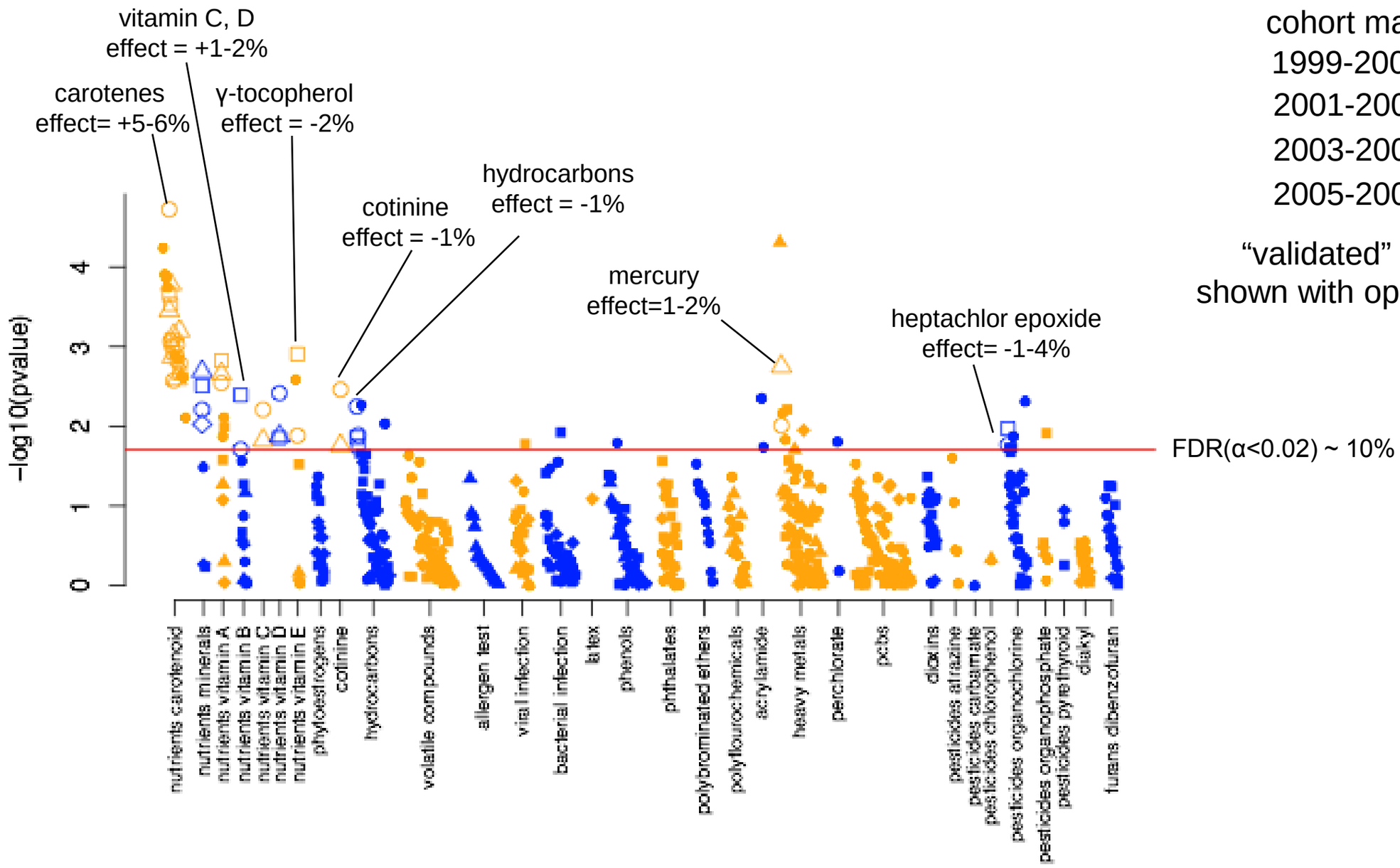


ORs adjusted:
BMI, SES, ethnicity, age, sex
Δ 1SD *or* vs. negative test

Patel CJ, Bhattacharya J, Butte AJ, (2010) An Environment-Wide Association Study (EWAS) on T2DM. *PLoS ONE* vol. 5(5)

What about other risk phenotypes?

EWAS on HDL-C



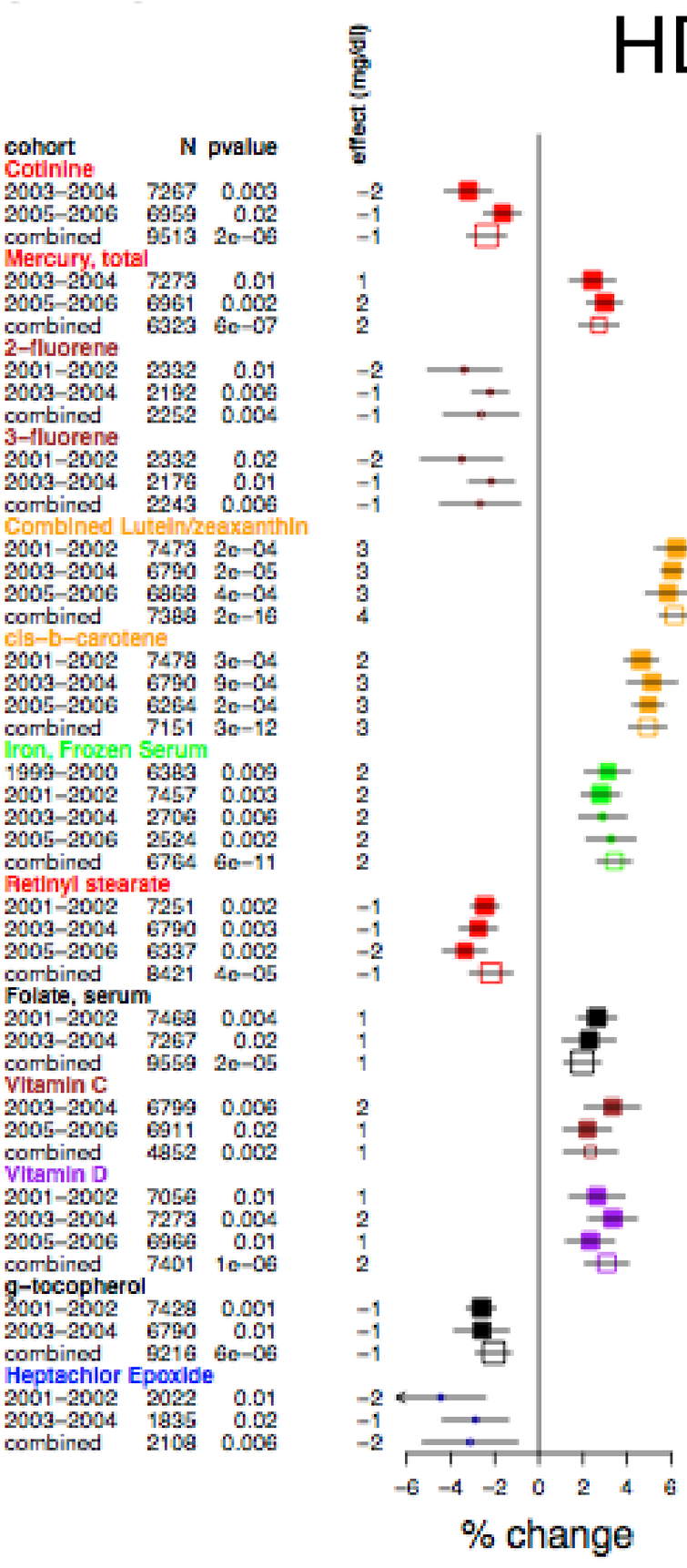
cohort markers
1999-2000 ◆
2001-2002 ■
2003-2004 ●
2005-2006 ▲

"validated" factors:
shown with open markers

adjusted for:
BMI, SES, ethnicity, age, age^2, sex
effect= % Δ for Δ 1SD *or* vs. negative test

Patel CJ, Cullen MR, Ioannidis JAP, J, Butte AJ, (2010). Non-genetic associations and correlation globes for determinants of Lipid Levels: an EWAS. In review.

Effect Sizes For Validated Factors: HDL-C



single cohorts adjusted for:
BMI, SES, ethnicity, age, age^2, sex
effect= % Δ for Δ 1SD
FDR ~ 10%

'combined' adjusted for:
BMI, SES, ethnicity, age, age^2,
sex, blood pressure, cohort,
diabetes, waist circumference
effect= % Δ for Δ 1SD
FDR < 1%

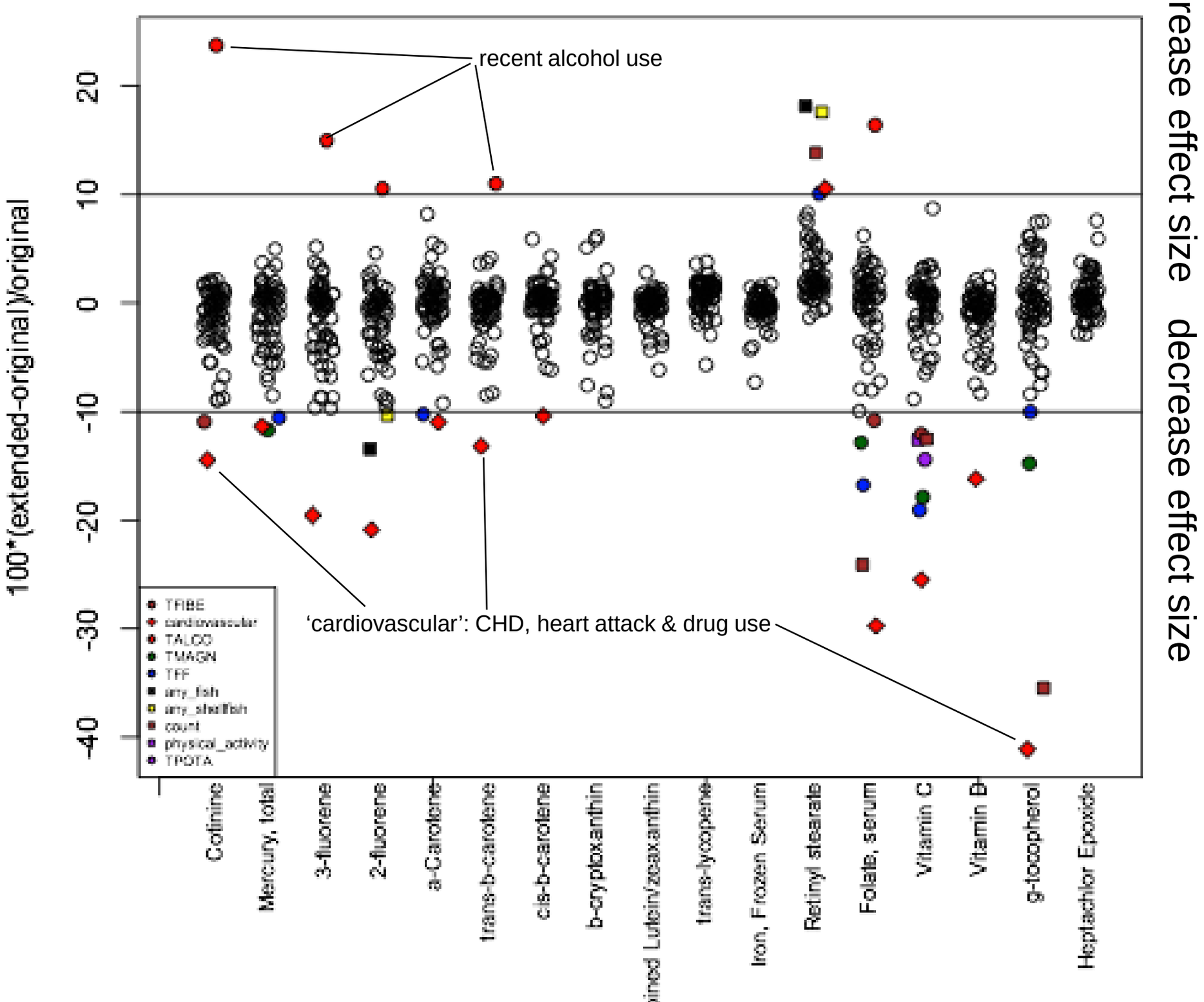
“Assessing” Bias from Self-Report Data

“source” of bias	examples
disease status	diabetes, CHD, heart attack
drug use	metformin, statins
supplement use	count of total supplements
physical activity	metabolic equivalents
recent food intake	total nutrients

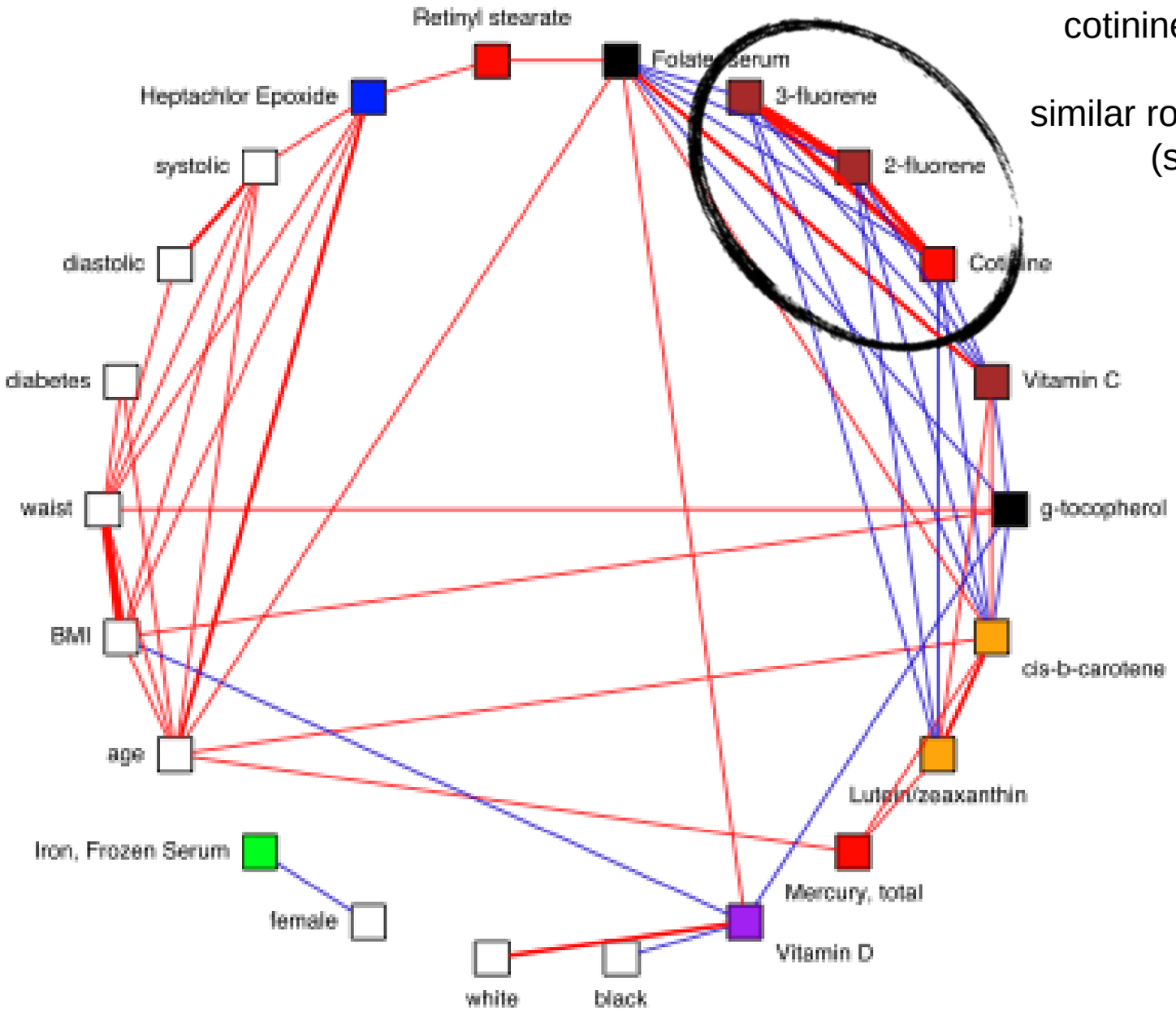
total:62

self-report data are sequentially added to final model
effect sizes for environmental variable are compared

Assessing Bias from Self-report data
Factors Validated for HDL-C



HDL-C “Correlation Globe”
Dependencies of Validated Factors



cotinine, 2-,3-fluorene
ρ > 0.6
similar route of exposure?
(smoking?)

red: positive correlation
blue: negative correlation
all |ρ| > 0.2

What’s next?

Validation: longitudinal studies & model systems

Genome-wide by exposome-wide studies

Missing Heritability*: Genome meets Exposome

Variants ascertained from genome-wide studies
have described little disease variability

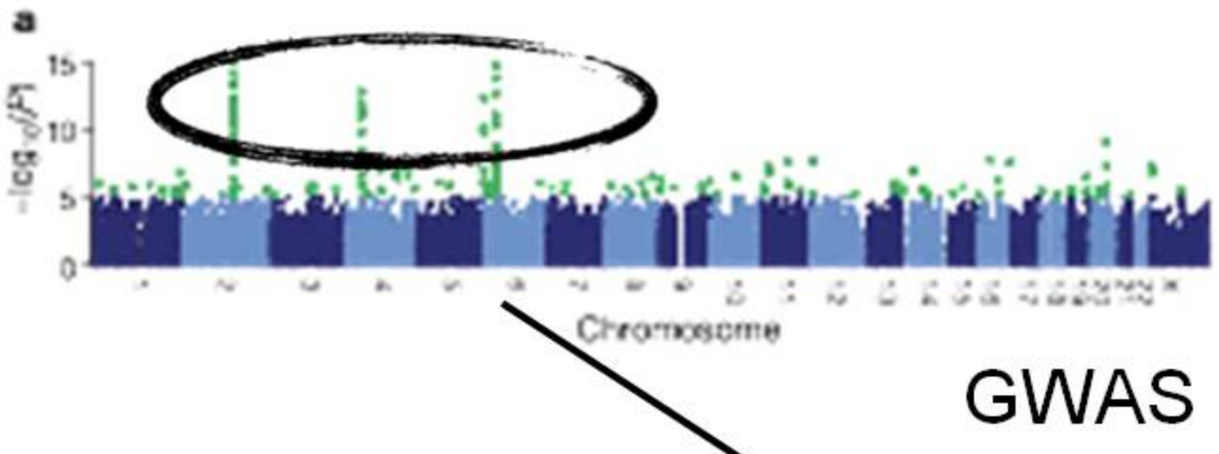
Type 2 Diabetes: 6% (18 loci)
HDL-Cholesterol: 5% (7 loci)

Considering genome in *combination* with exposome might give a better estimate of heritability

*Manolio T et al. Finding the missing heritability of complex diseases. Nature (2009) vol. 461 (7265) pp. 747

Do known disease-associated genetic variants interact with environmental factors found in EWAS?

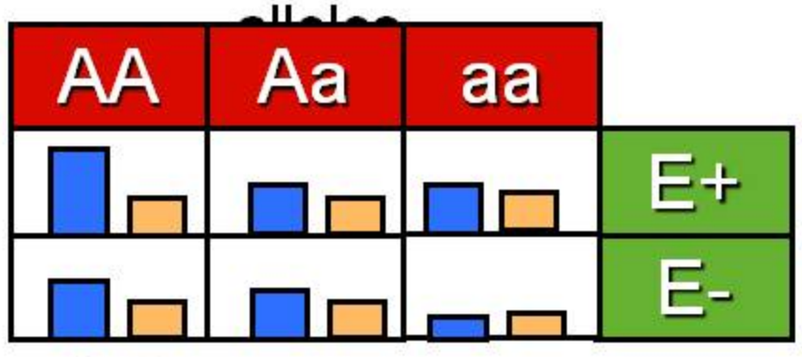
Example: Type 2 Diabetes

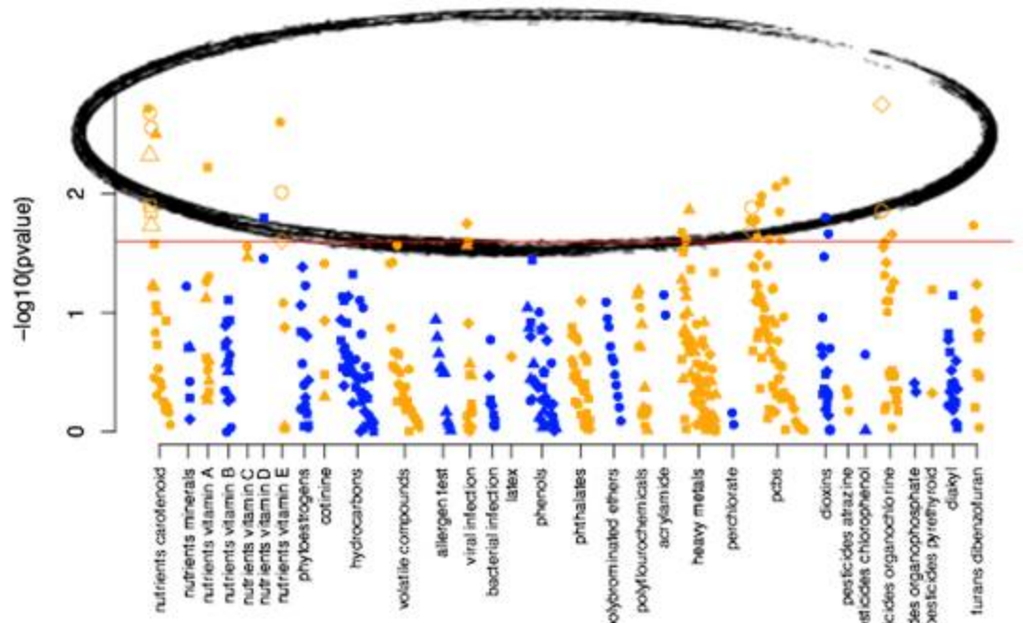


GWAS

TCF7L2 rs7903146

number of risk alleles





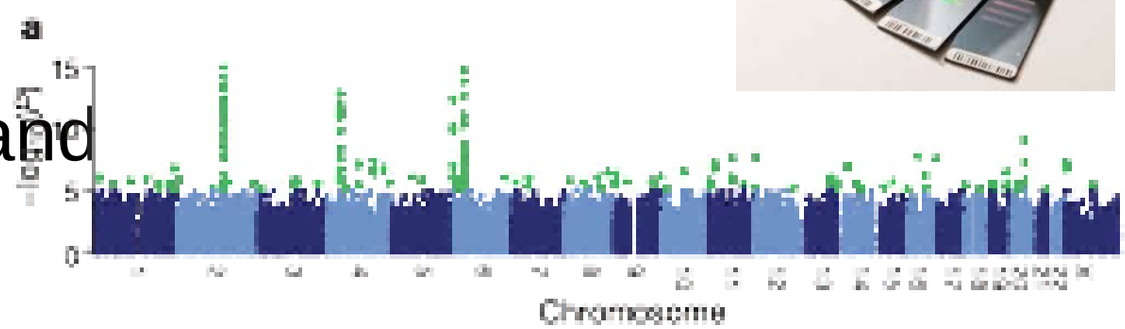
EWAS

[heptachlor]

To test:
NHANES: 1999-2003
~ 100 loci associated to metabolic related disease (T2D, BMI, lipids)
N ~ 3000

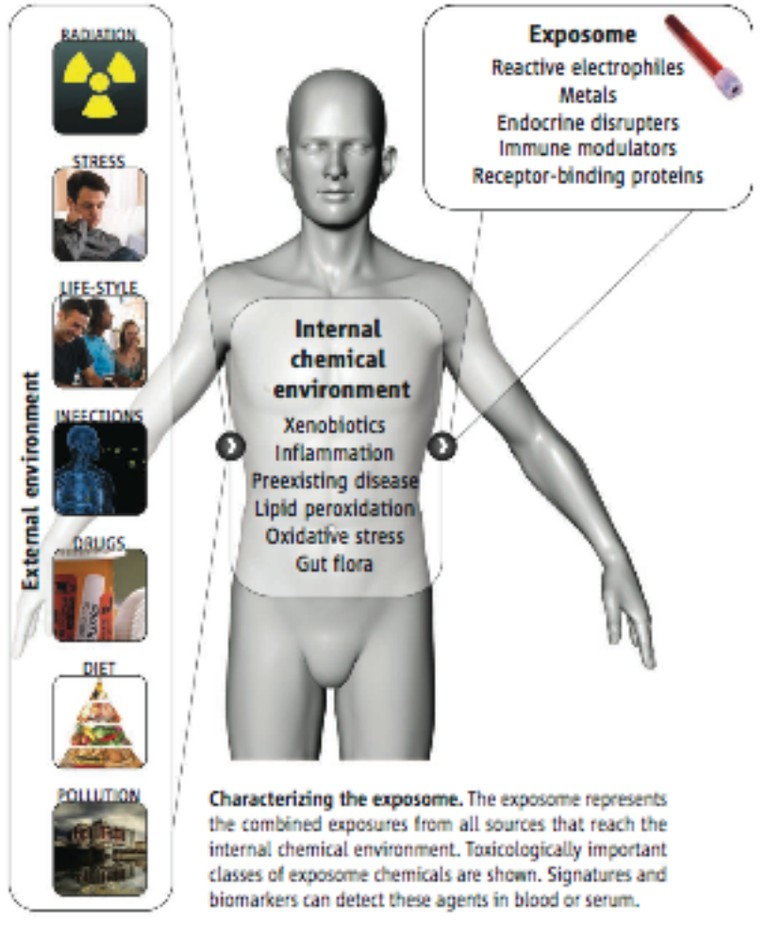
Conclusion and Discussion: EWAS and the “exposome”

- Comprehensive, transparent, and systematic



- *However*: reverse causality, confounding amplified

- Need to define, characterize, and make accessible the “exposome”



Acknowledgements & Thanks

Jay Bhattacharya

Mark Cullen

John Ioannidis

Atul Butte

CDC/NCHS/NHANES

Kris Thayer/NTP

Butte Laboratory

Biomedical Informatics Program

Stanford School of Pediatrics and Medicine

National Library of Medicine

Questions?

chirag.patel@stanford.edu

Butte Lab
Stanford Center for Biomedical Informatics Research